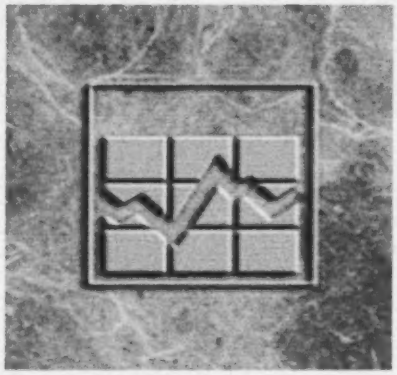
N° 89-648-X au catalogue
ISBN 978-0-660-05734-7

Série de documents de recherche de l'Étude longitudinale et internationale sur les adultes

Qualité du couplage rétrospectif des données : l'Étude longitudinale et internationale des adultes et les données fiscales sur les gains et le revenu



par James Hemeon

Date de diffusion : le 18 août 2016

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « Normes de service à la clientèle ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

.	indisponible pour toute période de référence
..	indisponible pour une période de référence précise
...	n'ayant pas lieu de figurer
0	zéro absolu ou valeur arrondie à zéro
0^s	valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
^{p}	provisoire
^{r}	révisé
x	confidentiel en vertu des dispositions de la *Loi sur la statistique*
^{E}	à utiliser avec prudence
F	trop peu fiable pour être publié
*	valeur significativement différente de l'estimation pour la catégorie de référence ($p<0,05$)

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2016

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'entente de licence ouverte de Statistique Canada.

Une version HTML **est aussi disponible.**

This publication is also available in English.

Table des matières

1. Introduction 4
2. Échantillon 4
3. Résultats 5
3.1 Taux de couplage entre 1982 et 2011 5
3.2 Examen du taux de couplage 7
3.3 Panels équilibrés 11
3.4 Comparaison des gains d'emploi tirés du T1FF et du fichier T4 11
3.5 Profils de gains selon l'âge et le sexe 12
4. Conclusions 14
Bibliographie 15
Annexe A. 16

1. Introduction

Statistique Canada recueille des données administratives à des fins statistiques depuis sa création en 1918 sous le nom de Bureau fédéral de la statistique. Le couplage des données des enquêtes avec les sources administratives est une pratique de plus en plus courante comme moyen de réduire le fardeau de réponse des répondants, de remplacer les questions de l'enquête par des données qui autrement pourraient donner lieu à un biais de rappel, et de recueillir des données qu'un répondant pourrait ne pas être à l'aise de divulguer au cours d'une interview d'enquête. De par sa nature, le couplage permet de réduire les coûts associés à la collecte de renseignements pour l'enquête. Depuis des décennies, les organismes statistiques de partout à travers le monde utilisent les données administratives en remplacement des questionnaires (Conseil économique et social, 2009).

Les données recueillies dans le cadre d'enquêtes de L'Étude longitudinale et internationale des adultes (ELIA) sont couplé avec les données fiscales et les données d'autres sources administratives pour chaque année de collecte. En outre, l'ELIA effectue un couplage rétrospectif avec les fichiers fiscaux précédant la première année de cueillette de l'étude, ce qui permet d'obtenir une quantité de données longitudinales importante sans alourdir le fardeau du répondant et sans coût de collecte supplémentaire. Même si les données socioéconomiques transversales, qui représentent un moment dans le temps, peuvent se révéler extrêmement utiles, la disponibilité d'un ensemble de données longitudinal de grande qualité comme l'ELIA permet de réaliser des analyses de tendances sur toute la durée de vie des gens, ce qui procure un éclairage plus complet lorsque vient le temps de prendre des décisions en matière de politiques publiques.

La présente étude a pour but d'explorer la qualité des données de couplage rétrospectif de l'ELIA. En effet, les avantages des données de couplage rétrospectif ont été présentés par Heisz et al (2013). Ces derniers ont utilisé les données d'une étude pilote pour analyser les taux de couplage des données fiscales et la précision des données. La présente étude se fonde sur certaines de ces mêmes méthodes et élargit ses résultats en les appliquant à l'ensemble de données de l'ELIA. Plus précisément, elle analyse le taux de couplage, sa décroissance au fur et à mesure que les données administratives remontent dans le temps ainsi que le potentiel des données de couplage rétrospectif dans l'analyse de phénomènes qui nécessitent des séries de données longitudinales.

2. Échantillon

L'ELIA est une enquête-échantillon employant une démarche stratifiée à plusieurs étapes et à plusieurs phases. L'échantillon a été composé en 2011 par la sélection de logements tirés des données du recensement de la population du Canada de 2011. Il s'agit donc d'un échantillon représentatif de la population à ce moment. Les premières interviews dans le cadre de l'ELIA se sont tenues à la fin de 2011 et au début de 2012. Par souci de simplicité, on utilisera le terme ELIA 2012 pour faire référence à cette première vague de collecte. C'est donc la base de données ELIA 2012 qui sera employée pour la présente étude. L'échantillon renferme des logements des dix provinces canadiennes, mais ne comprend pas de membres réguliers des Forces canadiennes, de personnes vivant en établissement ni de personnes vivant dans des réserves et autres établissements autochtones des provinces. Les données donnent des renseignements sur la démographie des répondants, la composition des familles et des ménages, les compétences en lecture, en calcul et en résolution de problèmes, la scolarité et la formation, la santé, le revenu et la richesse ainsi que sur la participation au marché du travail (Statistique Canada, 2014).

La base de données renferme les données de 23 926 répondants âgés de 15 ans et plus. Le fichier contient également 2 943 non-répondants et 5 264 enfants non répondants (de moins de 15 ans). Pendant les interviews de l'ELIA, les répondants ont été informés que les données recueillies dans le cadre de l'enquête seront reliées à des sources administratives, une pratique que l'on appelle « remplacement éclairé ».

À la suite de la cueillette, le numéro d'assurance sociale (NAS) des répondants a été extrait des bases de données fiscales de 2010 et 2011 à partir de leurs prénom, nom, date de naissance, sexe, état matrimonial, adresse et code postal (on ne demande pas directement aux répondants leur numéro d'assurance sociale). En cas de non-correspondance directe, on a couplé le NAS en ayant recours à des appariements établis selon un calcul de probabilités basé sur les variables auxiliaires citées précédemment.

Une fois le NAS identifié, les données de l'ELIA ont été couplées avec différents dossiers fiscaux des particuliers : i) le Fichier sur la famille T1 (T1FF), ii) l'état et le sommaire de rémunération payée par les employeurs (fichier T4) et iii) le fichier sur les Régimes de Pension au Canada[1]. Deux types de couplages ont été réalisés : i) un couplage annuel (renouvelable à chaque nouvelle vague de l'enquête) et ii) un couplage rétrospectif des données fiscales portant sur toutes les données remontant jusqu'à 1982 (Fichier sur la famille T1) ou 2000 (fichier T4, fichier sur les Régimes de Pension au Canada).

3. Résultats

3.1 Taux de couplage entre 1982 et 2011

Le couplage d'un enregistrement échoue lorsqu'il n'est pas possible de déterminer une clé d'appariement ou lorsque la clé d'appariement ne trouve pas de correspondance dans le fichier de données administratives.

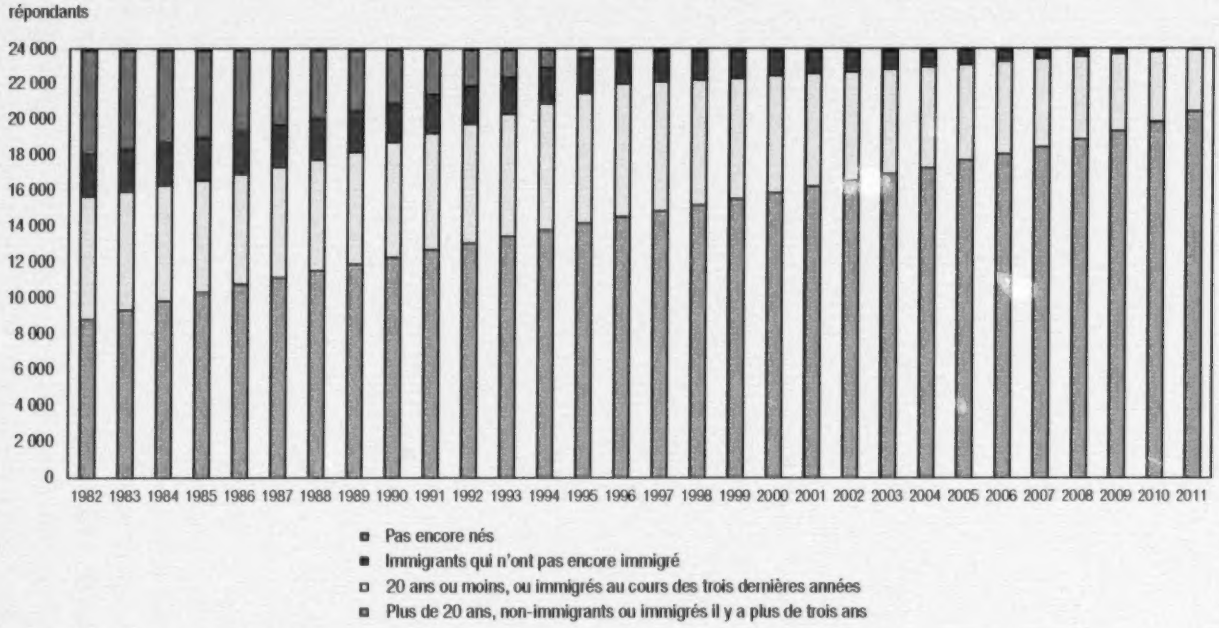
Comme nous l'avons souligné précédemment, les codes de NAS ont été établis à partir des bases de données fiscales de 2010 et 2011. Or, si aucun NAS n'a été trouvé au cours du processus de couplage, cela peut s'expliquer par le fait que le répondant n'a pas produit de déclaration de revenus personnelle durant ces deux années ou que le couplage probabiliste n'a pas permis d'établir de couplage adéquat. Ainsi, pour 7,5 % des répondants de l'ELIA, la tentative de couplage n'a pas permis de trouver un NAS. De ces répondants, 55,1 % avaient 17 ans et moins et 64,9 % avaient 20 ans et moins. On peut donc en conclure que, dans la majorité de ces cas, les répondants étaient jeunes et n'avaient pas besoin de produire une déclaration de revenus.

Dans le couplage rétrospectif de données, un problème supplémentaire se pose. En effet, le NAS d'une personne peut avoir changé à travers le temps. Si le NAS d'un répondant n'est pas constant dans le temps, le couplage échoue lorsqu'il n'est plus possible de retrouver le NAS. Même si le numéro d'assurance sociale (NAS) constitue une clé d'appariement relativement stable, dans certains cas, celui-ci peut changer à travers le temps. Par exemple, on attribue aux immigrants un NAS temporaire à leur arrivée au Canada avant de leur attribuer un NAS permanent. Si, pour cette raison, un NAS est introuvable, il peut manquer des informations dans l'ensemble de données longitudinal étant donné que le répondant peut avoir produit une déclaration de revenus au cours des premières années qu'il n'a pas été possible de lui associer.

En outre, le couplage rétrospectif pose un autre problème. En effet, lorsqu'on remonte dans les fichiers administratifs, il est possible qu'une personne ne se retrouve plus dans les fichiers parce qu'elle était trop jeune pour produire une déclaration ou, s'il s'agit d'un immigrant, parce qu'elle vivait peut-être dans un autre pays ou n'avait pas pris l'habitude de remplir de déclaration de revenu en utilisant un NAS permanent au Canada. L'échantillon de l'ELIA renferme des répondants pouvant n'être âgés que de 15 ans (en date de 2011), alors que le couplage rétrospectif avec les données du T1FF englobe 30 ans de données fiscales dans la publication de l'ELIA de 2012. Dans ce contexte, les données fiscales disponibles précèdent l'année de naissance de certains répondants de l'ELIA de même que l'année d'immigration au Canada de certains autres. Ainsi, en 1982, 24,4 % des répondants de l'ELIA n'étaient pas encore nés, 10,0 % n'avaient pas encore immigré au Canada et 28,6 % étaient âgés de 20 ans et moins ou avaient immigré au Canada au cours des trois dernières années. À partir de 1997, tous les répondants de l'ELIA étaient nés, 7,5 % n'avaient pas encore immigré au Canada et 30,2 % étaient âgés de 20 ans et moins ou avaient immigré au Canada au cours des trois dernières années (figure 3.1-1). Voilà pourquoi il est peu probable, voire impossible d'établir un couplage pour un sous-ensemble de l'échantillon de l'ELIA durant certaines années, ce qui indique qu'on peut s'attendre à ce que les taux de couplage rétrospectif déclinent au fur et à mesure qu'on remonte dans le temps.

1. L'Enquête sur les Régimes de Pension au Canada est un recensement annuel complet des régimes de pension agréés au Canada. D'autres couplages sont prévus pour les cycles futurs.

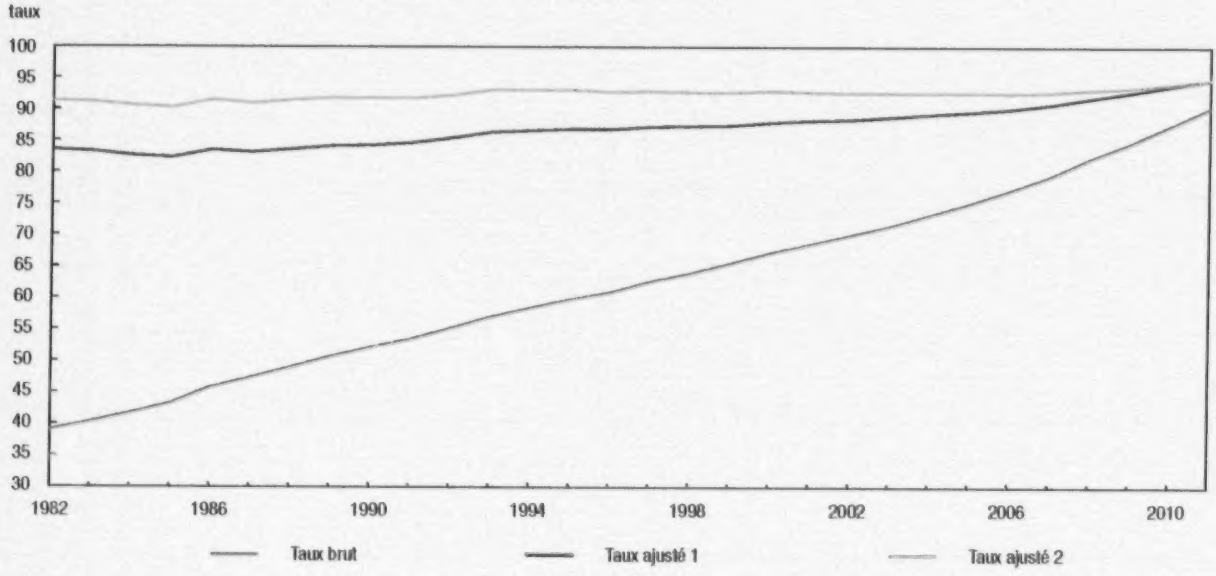
Figure 3.1-1
Caractéristiques de l'échantillon de l'ELIA par année, 1982 à 2011



Source : Étude longitudinale et internationale des adultes (2012).

Un taux de couplage a été calculé entre les répondants de l'enquête et les fichiers de données T1FF pour toutes les années de 1982 à 2011 afin d'en établir la baisse en remontant dans le temps. Trois différents taux de couplage ont été calculés : 1) un taux brut fondé sur l'ensemble de l'échantillon disponible, 2) un taux ajusté fondé sur un échantillon excluant les répondants âgés de 20 ans et moins durant une année d'imposition donnée et 3) un deuxième taux ajusté fondé sur un échantillon excluant les répondants âgés de 20 ans et moins ainsi que ceux ayant reçu leur statut d'immigrant au cours des trois années précédant une année d'imposition donnée (figure 3.1-2). Ces taux incluent les répondants pour lesquels aucun NAS n'a été trouvé. L'ajustement fondé sur l'âge repose sur le fait que ce groupe est moins susceptible de produire une déclaration de revenus durant une année donnée. L'ajustement fondé sur le statut d'immigrant repose sur le fait qu'il est peu probable que ce groupe ait produit une déclaration de revenus canadienne dans les années précédant leur immigration, et les trois années ont été choisies pour donner aux répondants immigrants le temps de prendre des habitudes de production de déclaration de revenus en utilisant un NAS permanent.

Figure 3.1-2
Taux de couplage de l'ELIA avec le Fichier sur la famille T1 (T1FF), de 1982 à 2011



Source : ELIA (2012) et données couplées avec le T1FF (1982 à 2011).

Comme l'illustre la figure 3.1-2, les résultats obtenus montrent que le taux de couplage diminue lorsqu'on recule dans le temps, peu importe l'échantillon sur lequel le calcul est fondé. Cependant, on note la plus importante diminution du taux de couplage lorsque le calcul est fondé sur l'échantillon sans restriction. Le taux chute alors considérablement (1,76 % par année, en moyenne), passant de 90,3 % en 2011 à seulement 39,1 % en 1982. Lorsqu'on exclut les répondants de moins de 20 ans dans une année d'imposition donnée, la baisse du taux est alors beaucoup moins importante (0,39 % en moyenne), passant de 94,8 % en 2011 à 83,6 % en 1982, et se maintenant à plus de 82 % durant toutes les années. En outre, lorsqu'on exclut les répondants qui ont immigré au Canada au cours des trois années précédant une année d'imposition donnée, on note que la diminution du taux de couplage est faible (0,12 % par année en moyenne), passant de 94,8 % en 2011 à 91,3 % en 1982, et se maintenant à plus de 90 % durant toutes les années. En d'autres termes, lorsqu'on limite l'échantillon à la population qui est susceptible de produire une déclaration de revenus et dont le NAS est constant, le taux de couplage demeure élevé durant toutes les années concernées.

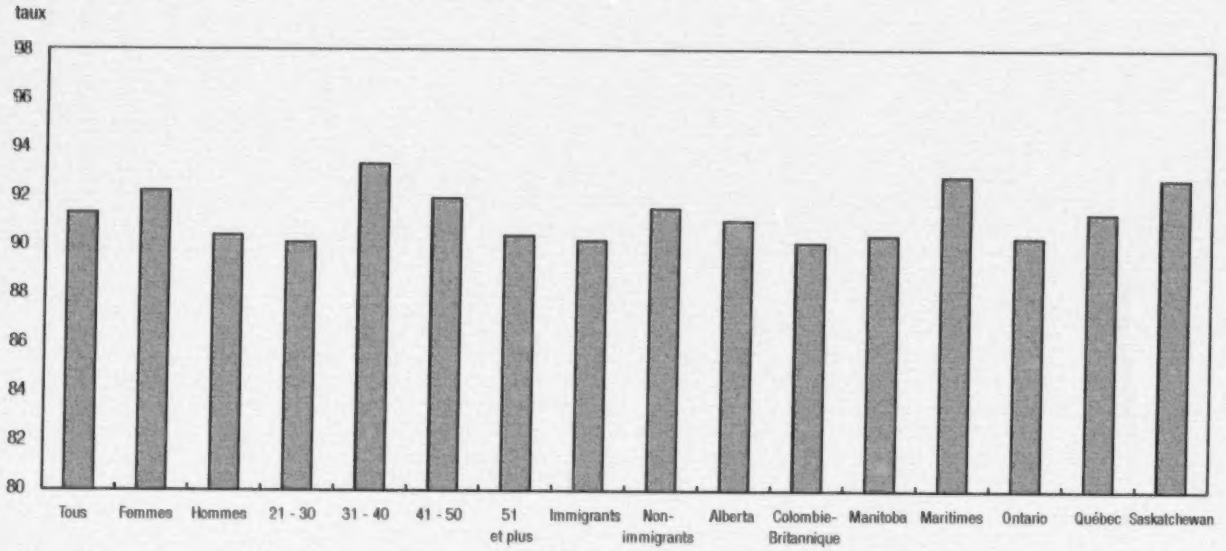
Dans l'échantillon de l'ELIA, 8,5 % des répondants n'ont pu être couplés aux données fiscales des années comprises entre 1982 et 2011. Ce groupe se compose de répondants pour lesquels on n'a pas trouvé de NAS et de ceux qui n'ont pas donné leur consentement au couplage[2].

3.2 Examen du taux de couplage

Afin d'analyser si les données de couplage étaient représentatives de l'échantillon, on a calculé plusieurs taux de couplage pour les répondants âgés de plus de 20 ans dans une année d'imposition donnée et qui n'étaient pas immigrants ou qui avaient immigré au cours des trois années précédant l'année d'imposition donnée (Taux ajusté 2, figure 3.1-2). En plus du taux de couplage global Ajusté 2, on a calculé des taux pour les sous-échantillons par âge (dans une année d'imposition donnée), sexe, statut d'immigrant et province de résidence (en date de 2011) pour les années d'imposition 1982, 1996 et 2011. En raison de la légère diminution du taux de couplage en 1985, comme l'illustre la figure 3.1-2, l'année 1985 a été incluse également (figures 3.2-1 à 3.2-4). Il est à noter que ces taux se fondent sur des fréquences non pondérées, car la présente étude a pour but d'analyser la qualité du couplage et non la représentativité de la population. Consultez le tableau 3.2-5 pour connaître le nombre total d'observations couplées en 1982, 1985, 1996 et 2011.

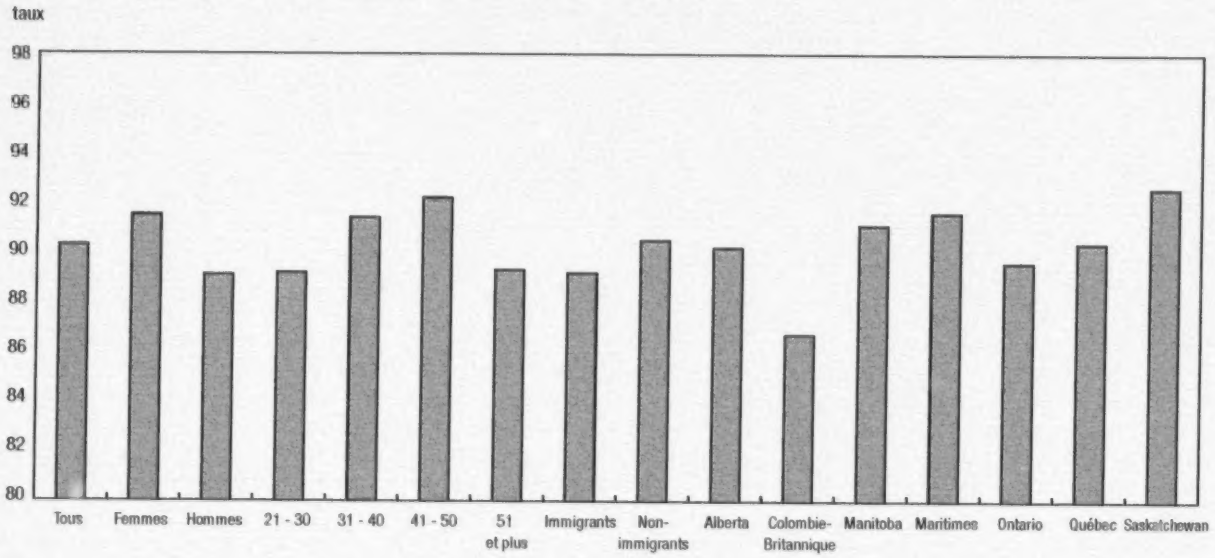
2. Conformément à la politique de Statistique Canada sur le couplage d'enregistrements, les répondants peuvent refuser le couplage d'enregistrements. Ce refus concernait un petit nombre de répondants de l'ELIA.

Figure 3.2-1
Taux de couplage ajusté 2 du sous-groupe démographique de l'ELIA, 1982



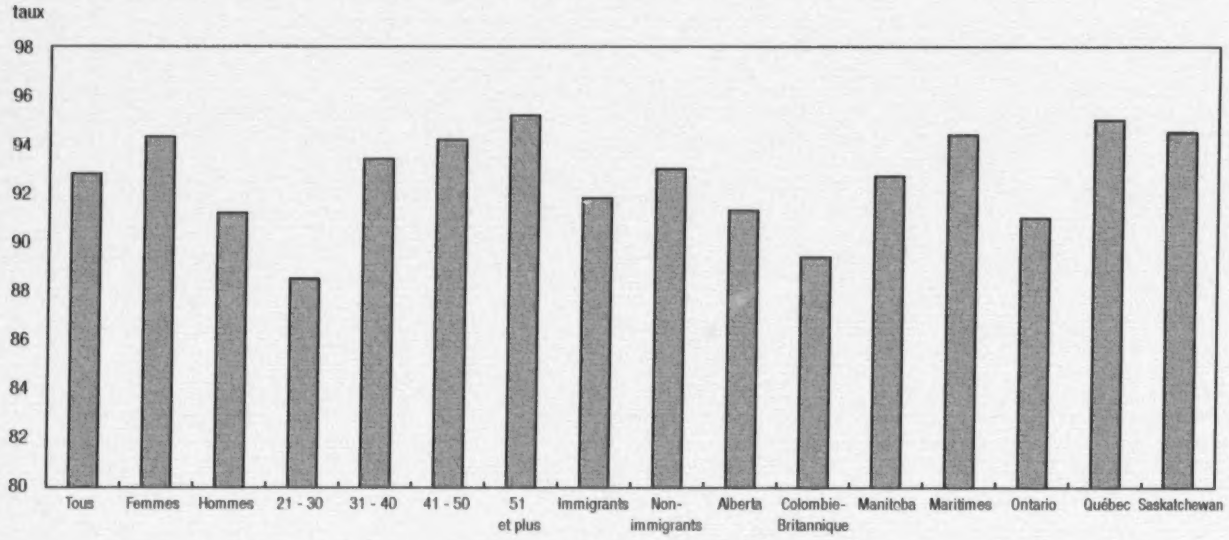
Source : ELIA (2012) et données couplées avec le T1FF (1982).

Figure 3.2-2
Taux de couplage ajusté 2 du sous-groupe démographique de l'ELIA, 1985



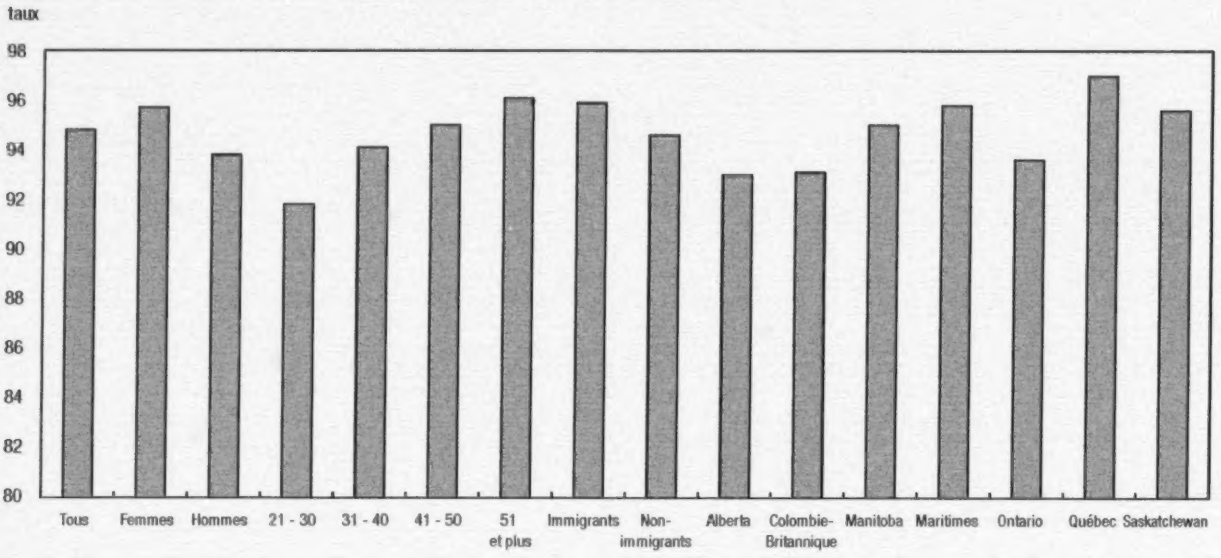
Source : ELIA (2012) et données couplées avec le T1FF (1985).

Figure 3.2-3
Taux de couplage ajusté 2 du sous-groupe démographique de l'ELIA, 1996



Source : ELIA (2012) et données couplées avec le T1FF (1996).

Figure 3.2-4
Taux de couplage ajusté 2 du sous-groupe démographique de l'ELIA, 2011



Source : ELIA (2012) et données couplées avec le T1FF (2011).

Les taux de couplage globaux pour 1982, 1985, 1996 et 2011 sont respectivement de 91,3 %, 90,3 %, 92,8 % et 94,8 %.

Le taux de couplage des hommes est légèrement inférieur à celui des femmes pour toutes les années, soit 95,7 % de couplage chez les femmes en 2011 par rapport à 93,8 % chez les hommes.

Les résultats montrent qu'en général, le taux de couplage augmente avec l'âge, ce qui concorde avec les résultats obtenus par Li et al (2006). Cependant, le taux des jeunes répondants, même s'il est inférieur, reste raisonnablement élevé à 91,8 %. De tous les groupes d'âges en 2011, ce sont les répondants âgés de 51 ans et plus au cours d'une année d'imposition donnée qui présentent le taux de couplage le plus élevé. Toutefois, ce taux est inférieur au cours des premières années étant donné que le nombre de répondants dans ce groupe d'âge pour une année d'imposition donnée diminue radicalement, passant de 8 879 en 2011 à 544 en 1982. Cette situation est prévisible, car les répondants du groupe d'âge de 51 ans et plus en 1982 avaient 81 ans et plus au moment de l'interview de l'ELIA. Par le fait même, ils étaient moins susceptibles de produire une déclaration de revenus en 2011 ou 2010 (ce qui diminuait les chances de trouver un NAS pour le couplage d'enregistrements).

Les immigrants présentent un taux de couplage supérieur à celui des non-immigrants en 2011. Généralement, les répondants qui résidaient en Ontario et en Colombie-Britannique en 2011 présentent un taux de couplage légèrement inférieur par rapport aux autres provinces. On constate cette situation plus particulièrement en 1985 où le taux était de 86,7 % pour la Colombie-Britannique, soit une baisse de 3,5 % par rapport à 1982. Les raisons de ce recul restent obscures. L'analyse des 312 répondants pour lesquels on avait un couplage d'enregistrements en 1982, mais pas en 1985, n'a pas permis d'établir de tendance en raison de l'âge, du sexe ou du statut d'immigrant.

Tableau 3.2-5
Observations du couplage d'enregistrements d'un sous-groupe de l'ELIA (1982, 1985, 1996, 2011)

Catégorie	1982	1985	1996	2011
Tous les participants	8 068	9 334	13 514	19 403
Femmes	4 153	4 878	7 220	10 247
Hommes	3 915	4 456	6 294	9 156
21-30	3 991	4 168	2 911	3 040
31-40	2 460	3 102	4 630	3 043
41-50	1 073	1 246	3 581	4 441
51 et plus	544	818	2 392	8 879
Immigrants	889	1 038	1 829	3 558
Non-immigrants	7 179	8 296	11 685	15 845
Résidents de l'Alberta 2011	735	860	1 313	2 056
Résidents de la Colombie-Britannique 2011	860	948	1 396	2 129
Résidents du Manitoba 2011	576	679	963	1 373
Résidents des Maritimes 2011	2 043	2 379	3 278	4 138
Résidents de l'Ontario 2011	1 716	1 963	2 930	4 493
Résidents du Québec 2011	1 593	1 869	2 734	3 920
Résidents de la Saskatchewan 2011	545	636	900	1 294

Source : ELIA (2012) et données couplées du T1FF (1982, 1985, 1996, 2011).

3.3 Panels équilibrés

Pour constituer un ensemble de données longitudinales, il faut disposer de données sur l'échantillon échelonnées sur une période donnée. Toute rupture dans la continuité des données risque d'en limiter l'utilité pour les chercheurs dans certaines conditions. On dit alors d'un ensemble de données longitudinal qu'il constitue un « panel équilibré » lorsque toutes les observations (répondants) qui le composent sont présentes à toutes les périodes (dans le cas de l'ELIA, chaque année). Dans le cas d'un couplage rétrospectif, pour que le panel soit équilibré, il faut établir un couplage avec l'enregistrement fiscal de chaque année.

Si, par exemple, un chercheur a besoin d'un panel équilibré formé à partir des répondants susceptibles de produire une déclaration de revenus et d'avoir un NAS constant dans le temps (Taux ajusté 2, figure 3.1-2), on pourrait alors constituer un panel de 30 ans avec un taux de couplage de 74,3 % regroupant 6 564 répondants (pour les années 1982 à 2011). Si un chercheur avait besoin d'un panel de 25 ans, on pourrait le former avec un taux de couplage de 78,1 % (1987-2011), et il contiendrait 8 735 répondants. On pourrait former un panel de 20 ans avec un taux de couplage de 82,1% (1992-2011) regroupant 10 733 répondants. On pourrait former un panel de 15 ans avec un taux de couplage de 84,5% (1997-2011) regroupant 12 579 répondants. On pourrait former un panel de 10 ans avec un taux de couplage de 86,7% (2002-2011) regroupant 14 371 répondants. Si un chercheur avait besoin d'un panel de seulement 5 ans, on pourrait le former avec un taux de couplage de 89,7% (2007-2011) regroupant 16 568 répondants (Annexe A). Ainsi, il est possible d'utiliser l'ELIA pour créer de longs panels équilibrés d'une taille suffisante pour mener de nombreuses analyses.

3.4 Comparaison des gains d'emploi tirés du T1FF et du fichier T4

L'une des façons de vérifier la fiabilité des fichiers administratifs consiste à en comparer les données avec celles des fichiers administratifs d'une autre source.

Les montants de gains d'emploi, présents autant dans les fichiers de données T1FF que dans les fichiers de données T4, ont été comparés pour la période allant de 2000[3] à 2011. Les résultats obtenus montrent que la majorité des enregistrements, soit environ 97 % chaque année, présentent une situation de gains qui concorde aussi bien avec le Fichier sur la famille T1 (T1FF) qu'avec le fichier de l'employeur (T4) (Tableau 3.4-1). En d'autres termes, ce n'est que dans 3 % (environ) des cas que les gains sont présents dans un fichier et absents de l'autre. Dans approximativement 71 % des cas, les gains sont présents à la fois dans le T1FF et dans le fichier T4. Pour 26 % des répondants, on relevait des gains de 0 $ dans le T1FF et aucun renseignement dans le fichier T4. Le nombre de cas où on ne retrouve aucune donnée dans le T1FF et des gains de 0 $ dans le fichier T4 est négligeable.

Tableau 3.4-1
Source de gains, T1FF et fichier T4, 2000 à 2011

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
						%						
Gains - des deux sources (T1FF et T4 >= 0 $)	72,7	72,2	71,9	71,6	71,0	71,3	71,0	70,9	70,5	69,7	69,7	70,1
Gains d'une seule source (soit T1FF, soit T4 > 0 $)	3,1	3,2	3,1	3,3	3,5	3,2	3,6	3,6	3,5	2,9	2,3	1,3
Aucun gain - une seule source (T1FF = 0 $, rien dans T4)	24,2	24,6	25,0	25,2	25,4	25,5	25,4	25,5	26,0	27,4	28,1	28,6
Aucun gain - une seule source (rien dans T1FF, T4 = 0 $)	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0

Source : ELIA (2012) et données couplées du T1FF et du fichier T4 (2000-2011).

Dans la grande majorité des cas, on retrouve des gains dans les fichiers T1FF et T4, ou des gains de 0 $ dans le T1FF et aucune donnée dans le fichier T4 (ce qui indique une concordance entre les deux fichiers). Lorsque des valeurs de gains figurent dans les deux fichiers (T1FF et T4), dans près de 95 % des cas, on relève au plus une différence d'un dollar entre les deux sources de données (Tableau 3.4-2). Il est à noter que les gains du fichier T4 tiennent compte des cents dans les montants alors que ce n'est pas le cas du T1FF. Dans 98 % des cas, environ, on note une différence ne dépassant pas mille dollars entre les sources de données.

3. Les renseignements des fichiers T4 n'étaient pas disponibles pour les années antérieures à l'an 2000.

Tableau 3.4-2
Différence entre les gains d'emploi déclarés selon le T1FF et le fichier T4, 2000 à 2011

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
						%						
0,01 $ à 1,00 $	92,7	95,0	94,2	94,6	94,5	94,4	94,2	94,7	95,1	95,7	96,1	96,5
0,01 $ à 100,00 $	2,2	1,9	1,7	1,6	1,6	1,8	1,8	1,4	1,3	1,0	0,8	1,0
100,01 $ à 1 000 $	1,8	1,5	1,8	1,9	1,8	1,7	1,8	1,9	1,6	1,5	1,3	1,1
T1FF < T4 de plus de 1 000 $	2,6	0,9	1,5	1,2	1,3	1,3	1,5	1,3	1,3	1,2	1,2	0,9
T1FF > T4 de plus de 1 000 $	0,7	0,7	0,8	0,6	0,9	0,9	0,7	0,7	0,8	0,6	0,6	0,5

Source : ELIA (2012) et données couplées du T1FF et du fichier T4 (2000-2011).

De 2000 à 2011, la différence entre les gains d'emploi médians calculés à partir des deux sources de données s'élève en moyenne à 116 $ (Tableau 3.4-3). Lorsqu'ils sont présents dans les deux sources (T1FF et le fichier T4), les gains médians sont très semblables, ce qui indique que le couplage des données du T1FF est juste et que celui du fichier T4 est également présent comme on peut le prévoir.

Lorsque les données sont présentes uniquement dans une source, les gains médians sont considérablement inférieurs à ceux qu'on obtient lorsque les données se retrouvent dans les deux sources de données. Un examen plus approfondi de cette situation montre que la majeure partie des valeurs d'une seule source sont de 0 $ provenant du T1FF sans valeur correspondante dans le fichier T4. La majorité des gains d'une seule source supérieure à 0 $ proviennent de valeurs contenues dans le fichier T4, mais absentes du T1FF.

Tableau 3.4-3
Gains d'emploi médians[1] du T1FF et du fichier T4

	Deux sources			Une source, > 0 $			
	T1FF		T4	T1FF		T4	
Année	N	Médiane	Médiane	N	Médiane	N	Médiane
2000	12 000	31 979	32 584	84	7 248	430	6 294
2001	12 192	33 109	33 152	78	11 396	462	6 822
2002	12 377	32 517	32 634	67	11 545	461	8 273
2003	12 601	32 727	32 815	57	6 559	515	5 374
2004	12 808	32 814	32 851	141	26 926	494	5 940
2005	13 162	33 317	33 330	89	9 903	495	5 901
2006	13 523	33 541	33 653	98	11 574	588	5 311
2007	13 917	33 816	33 942	78	12 624	628	4 787
2008	14 287	33 983	34 052	123	12 612	589	5 564
2009	14 473	33 844	34 009	113	14 004	489	5 728
2010	14 853	33 322	33 410	94	10 724	390	8 912
2011	15 290	33 073	33 133	96	11 457	181	19 543

1. Gains exprimés en dollars constants de 2011.
Source : ELIA (2012) et données couplées du T1FF et du fichier T4 (2000-2011)

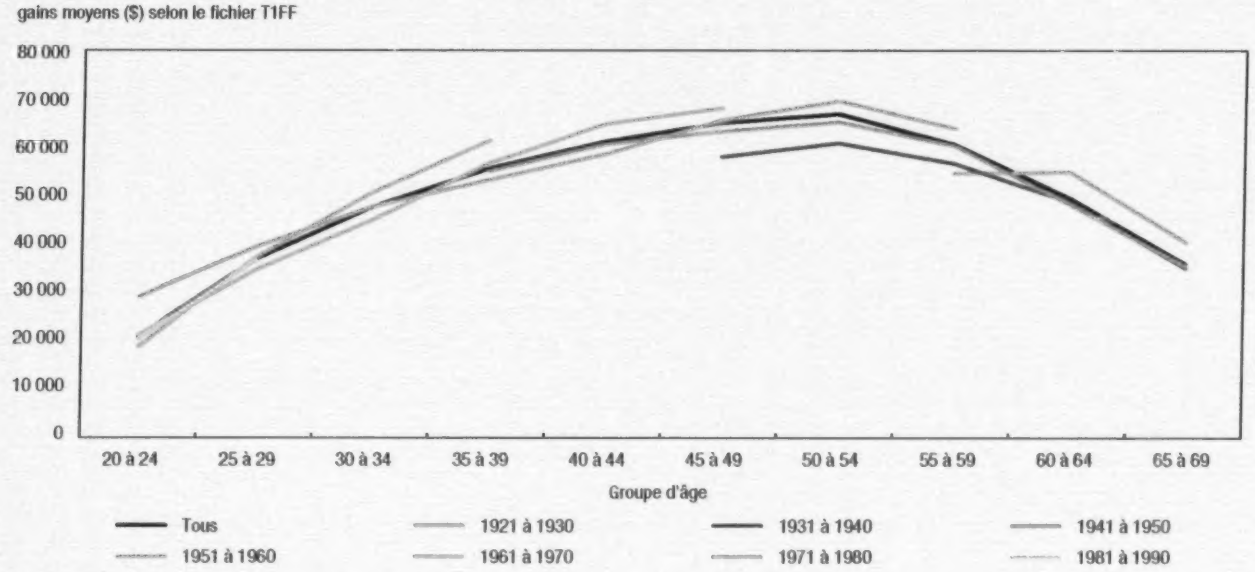
3.5 Profils de gains selon l'âge et le sexe

Afin de démontrer le potentiel des données couplées de l'ELIA pour créer de longues séries de données, nous avons créé un profil des gains en fonction de l'âge pour chaque sexe et pour différentes cohortes de dates de naissance.

En raison de la concordance des gains observée en comparant le Fichier sur la Famille T1 (T1FF) et les fichiers des entreprises (T4), nous avons retenu les gains contenus dans les T1FF, car ils permettaient de produire des

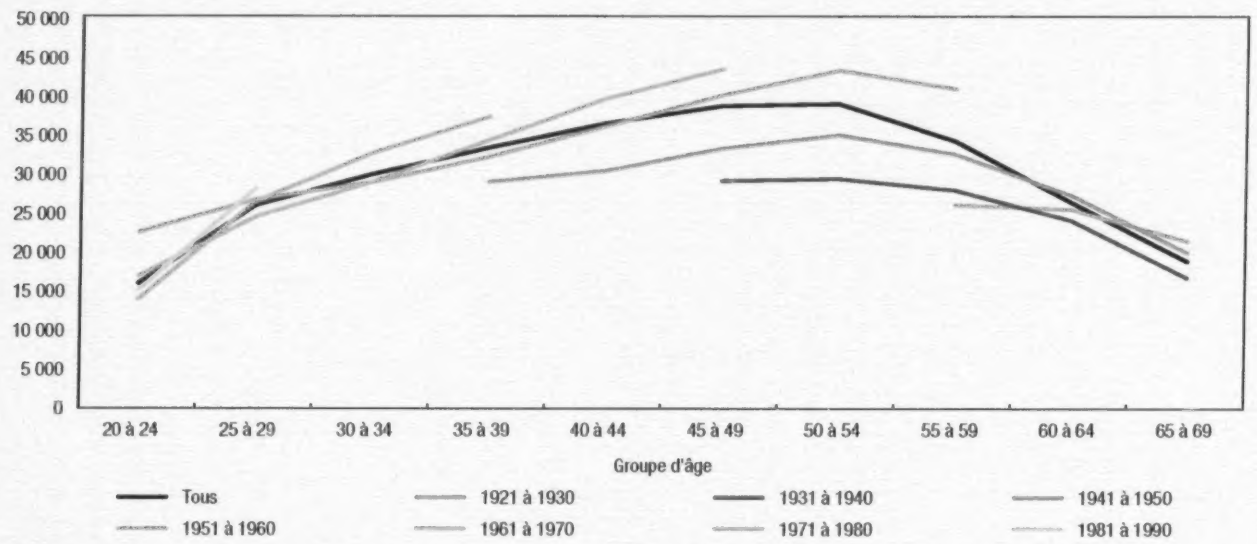
séries de données plus longues. Nous avons divisé l'échantillon en sept groupes d'âge de naissance par tranches de cinq ans pour lesquelles nous avons suivi les changements de gains d'emplois par âge.

Figure 3.5-1
Profils de gains des hommes par groupe d'âge et cohorte d'année de naissance



Source : ELIA (2012) et données couplées avec le T1FF de 1982 à 2011.

Figure 3.5-2
Profils de gains des femmes par groupe d'âge et cohorte d'année de naissance



Source : ELIA (2012) et données couplées avec le T1FF de 1982 à 2011.

Les profils des gains en fonction de l'âge par cohorte d'année de naissance montrent que les gains ont tendance à être plus faibles en début de carrière et croissent rapidement chez les travailleurs des cohortes récentes (p. ex., cohorte de 1971-1980 versus cohorte de 1961-1970). On note également que le maximum de gains a tendance à être supérieur chez les travailleurs des cohortes récentes (p. ex., cohorte de 1951-1960 versus cohorte de 1941-1950). Ces tendances concordent avec celles que l'on trouve dans la documentation à ce sujet (Vijay et al, 2014; Beach et Finnie, 2004).

Fait à noter, la progression du revenu gagné durant la carrière est plus importante chez les femmes dans chaque cohorte successive. Ainsi, le revenu des femmes de 50 à 54 ans de la cohorte de 1951-1960 a été 47 % supérieur à celui des femmes de la cohorte de 1931-1940 lorsque ces dernières avaient 50 à 54 ans. En comparaison, l'augmentation du revenu chez les hommes de 50 à 54 ans pour les cohortes respectives a été de 14 %. Ces résultats confirment les observations antérieures selon lesquelles l'augmentation des gains des femmes est supérieure à celle des hommes (Williams, 2010; Suh, 2010; Blau et Kahn, 2006).

4. Conclusions

La présente étude apporte une évaluation partielle de la qualité des données administratives de 1982 à 2011 couplées aux données de l'ELIA de 2012. Plus particulièrement, elle a consisté à analyser les taux de couplage, à comparer les données entre différentes sources administratives et à évaluer la pertinence d'utiliser les données de couplage pour analyser des phénomènes sélectionnés nécessitant des séries de données longitudinales.

Les taux de couplage avec les données administratives ont été étudiés sous différents angles. Les résultats obtenus indiquent que les taux de couplage d'enregistrements sont élevés, car plus de 90 % des répondants de l'ELIA âgés de 15 ans et plus ont pu être appariés en 2011. Les taux de couplage des années antérieures ont été élevés aussi, en particulier lorsqu'ils étaient calculés pour les répondants de 20 ans et plus et ceux qui avaient immigré au pays au moins trois ans avant l'année du couplage. Le taux de couplage demeure élevé parmi les sous-groupes démographiques clés. Toutefois, les utilisateurs des données doivent tenir compte du fait que le nombre d'observations rétrospectives peut être inférieur chez certains sous-groupes. Ainsi, les données de couplage des immigrants n'apparaîtront qu'à partir de l'année de leur arrivée au Canada.

Les résultats suggèrent également que les données obtenues au moyen de méthodes de couplage rétrospectif produisent des données cohérentes avec celles des différentes sources de données administratives. Elles peuvent donc être employées pour observer des phénomènes nécessitant le recours à des séries de données longitudinales de même que de longs ensembles de données de panel.

Étant donné que les données reposent sur un échantillon constitué en 2011, elles se prêtent mieux à des études destinées à décrire les histoires au cours de la vie d'une cohorte en particulier, plutôt qu'à établir un référencement transversal avec des années précises. Le couplage permet de réaliser des analyses rétrospectives des données sur les revenus qu'il n'aurait pas été possible de réaliser autrement sans les données de 30 années d'enquêtes, ou sans introduire un important biais de rappel. En outre, les prochaines publications de données seront couplées à d'autres années de données de l'ELIA, ce qui accroîtra le potentiel analytique de l'ensemble de données.

Bibliographie

Beach, C. et Finnie, R. (2004), "A Longitudinal Analysis of Earnings Change in Canada", Analytical Studies Branch. Research Paper, Statistics Canada.
http://www.statcan.gc.ca/pub/11f0019m/11f0019m2004227-eng.pdf

Blau, Francine D.; Kahn, Lawrence M. (2006) "The US gender pay gap in the 1990s: slowing convergence", IZA Discussion Papers, N° 2176
www.econstor.eu/dspace/bitstream/10419/34046/1/51436131X.pdf

Economic and Social Council. (2009) "Main Results Of The UNECE-UNSD Survey On The 2010 Round of Population and Housing Censuses", Economic Commission for Europe, Conference of European Statisticians. Twelfth Meeting, 28-30 October 2009.
http://unstats.un.org/unsd/censuskb20/Attachment459.aspx

Gill, Vijay, Knowles, James, Stewart-Patterson, David. (2014). "The Buck Stops Here: Trends in Income Equality Between Generations", Ottawa: The Conference Board of Canada.

Heisz, Andrew, Langevin, Manon, Randle, Jeffrey. (2013). "Historical data linkage of tax records on labour and income: The case of the Living in Canada Survey pilot". Statistics Canada Catalogue n° 89-648-X (2).
http://www.statcan.gc.ca/pub/89-648-x/89-648-x2013002-eng.htm

Li, Bing, Quan, Huge, Fond, Andrew, Lu, Mingshan. (2006) "Assessing record linkage between health care and Vital Statistics databases using deterministic methods", BioMed Central Health Services Research 2006, 6:48.
http://www.biomedcentral.com/1472-6963/6/48/

Sakshug, Joseph W., Couper, Mick P., Ofstedal, Mary B., Weir, David R. (2012) "Linking Survey and Administrative Records: Mechanisms of Consent", Sociological Methods & Research, 41(4) 535-569.
http://smr.sagepub.com/content/41/4/535.full.pdf

Statistics Canada. (2014). LISA Detailed information for 2014 (Wave 2).
http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5144

Suh, Jingyo. (2010) "Decomposition of the Change in the Gender Wage Gap", Research in Business and Economics Journal, 2-18.
http://www.aabri.com/manuscripts/08076.pdf

Williams, Cara. (2010) "Women in Canada: A Gender-based Statistical Report. Sixth Edition". Economic Well-being. Statistics Canada Catalogue n° 89-503-X. p. 32-33.
http://www.statcan.gc.ca/pub/89-503-x/2010001/article/11388-eng.pdf

Annexe A.

Panels équilibrés de l'ELIA

		5 ans	10 ans	15 ans	20 ans	25 ans	30 ans
1982	%	85,3 %	80,6 %	78,9 %	77,3 %	75,9 %	74,3 %
	N	7 531	7 123	6 966	6 831	6 700	6 564
1987	%	85,7 %	83,2 %	81,5 %	79,7 %	78,1 %	
	N	9 582	9 307	9 112	8 918	8 735	
1992	%	88,8 %	86,2 %	84,0 %	82,1 %		
	N	11 606	11 265	10 978	10 733		
1997	%	89,3 %	86,6 %	84,5 %			
	N	13 305	12 889	12 579			
2002	%	89,2 %	86,7 %				
	N	14 798	14 371				
2007	%	89,7 %					
	N	16 568					

Source : ELIA (2012) et données couplées avec le T1FF de 1982 à 2011.